УДК 004,7.056

В.В. Анищенко, Д.А. Вятченин, А.В. Доморацкий, В.К. Фисенко
Объединенный институт проблем информатики НАН Беларуси, г. Минск
Беларусь, 220012, г. Минск, ул. Сурганова, 6

Выявление аномального поведения системами обнаружения атак при интервально-значном представлении данных

V.V. Anishchenko, D.A. Viattchenin, A.V. Damaratski, V.K. Fisenko
United Institute of Informatics Problems
NAS of Belarus, c. Minsk
Belarus, 220012, c. Minsk, Surganov st., 6

Detecting Anomalous Behavior via Discovering Attacks Systems for Interval-Valued Data

В.В. Анищенко, Д.А. Вятченін, А.В. Доморацький, В.К. Фісенко
Об'єднаний інститут проблем інформатики НАН Білорусі, м. Мінськ
Білорусь, 220012, м. Мінськ, вул. Сурганова, 6

Вияв аномальної поведінки системами виявлення атак при інтервально-значному зображенні даних

В статье рассмотрен метод обнаружения аномального поведения пользователей распределенной компьютерной сети при интервально-значном представлении данных, основанный на построении устойчивой кластерной структуры с помощью эвристического метода возможностной кластеризации. Предложенный метод иллюстрируется результатами вычислительного эксперимента.
Ключевые слова: кластеризация, аномальные наблюдения, неопределенные данные.

A method of detecting anomalous user behavior in a distributed computational network for a case of interval-valued data is considered in the article. The method is based on constructing stable clustering structure using a heuristic method of possibilistic clustering. The proposed method is illustrated by the results of numerical experiment.
Key words: clustering, anomalous observations, uncertain data.

У статті розглянуто метод виявлення аномальної поведінки користувачів розподіленої комп'ютерної мережі при інтервально-значному зображенні даних, що заснований на побудові стійкої кластерної структури за допомогою евристичного методу можливісної кластеризації. Запропонований метод ілюструється результатами обчислювального експерименту.
Ключові слова: кластеризація, аномальні спостереження, невизначені дані.

Введение

Системы обнаружения атак на объекты информатизации давно применяются как один из необходимых рубежей защиты информационных систем. Системы обнаружения атак представляют собой программные или аппаратно-программные решения, которые автоматизируют процесс контроля событий, протекающих в компьютерной системе или сети, а также самостоятельно анализируют эти события в поисках признаков проблем

безопасности. Так как количество различных типов и способов организации несанкционированных проникновений в чужие сети значительно увеличилось за последние годы, системы обнаружения атак стали необходимым компонентом инфраструктуры безопасности большинства организаций.

Системы обнаружения атак условно делятся на два типа: системы обнаружения злоумышленного поведения и системы обнаружения аномального поведения. Системы обнаружения злоумышленного поведения основаны на информации о признаках, характеризующих поведение злоумышленника, тогда как работа систем обнаружения аномального поведения основана на информации о некоторых признаках, характеризующих допустимое поведение объекта наблюдения, где под допустимым поведением понимаются действия, выполняемые объектом и не противоречащие политике безопасности. Главным достоинством систем обнаружения аномального поведения является возможность генерирования системами указанного типа информации, которая может быть использована в системах обнаружения злоумышленного поведения, что, в свою очередь, открывает возможности создания гибридных систем обнаружения атак.

Наиболее распространенным видом реализации технологии обнаружения аномального поведения является применение различных статистических методов, в том числе кластерного анализа [1]. Следует также указать, что в работе [1] особо отмечается высокая эффективность методов кластеризации в задачах обнаружения аномальных наблюдений в исследуемой совокупности объектов. В последние годы особый интерес у исследователей вызывают методы нечеткой и возможностной кластеризации [2], отличительной чертой которых является не просто указание принадлежности того или иного объекта к определенному кластеру, но и степень, с которой данный объект принадлежит тому или иному таксону. Необходимо отметить, что подавляющее большинство алгоритмов нечеткой и возможностной кластеризации являются представителями так называемого оптимизационного направления. С другой стороны, в работах [3-7] предложен так называемый эвристический подход к решению задачи возможностной кластеризации, отличающийся от оптимизационного подхода устойчивостью результатов классификации.

В работе [8] предложена методология применения эвристических алгоритмов возможностной кластеризации при разработке систем обнаружения аномального поведения, которая основывается на выработке признакового пространства, описывающего нормальное поведение объектов информатизации, с последующим сбором статистической информации и обнаружением объектов, поведение которых отличается от допустимого. Недостатком предложенного в [8] подхода является описание поведения объектов информатизации в виде вектора некоторых количественных признаков, что является приемлемым для компьютерных систем, насчитывающих сравнительно небольшое число однотипных элементов, примером которых являются локальные вычислительные сети.

Целью данной работы является модификация предложенной в [8] методологии для случая распределенных вычислительных систем, поведение элементов которых может описываться векторами интервалов.

Представление данных о поведении элементов распределенной вычислительной сети

Распределенные вычислительные системы представляют собой совокупность значительно удаленных друг от друга отдельных ЭВМ и локальных сетей, представляющих собой вычислительные узлы. Распределенные вычислительные системы используются

для решения как наборов независимых задач, так и единой сложной задачи. На рис. 1 приведена упрощенная схема организации распределенной вычислительной сети.

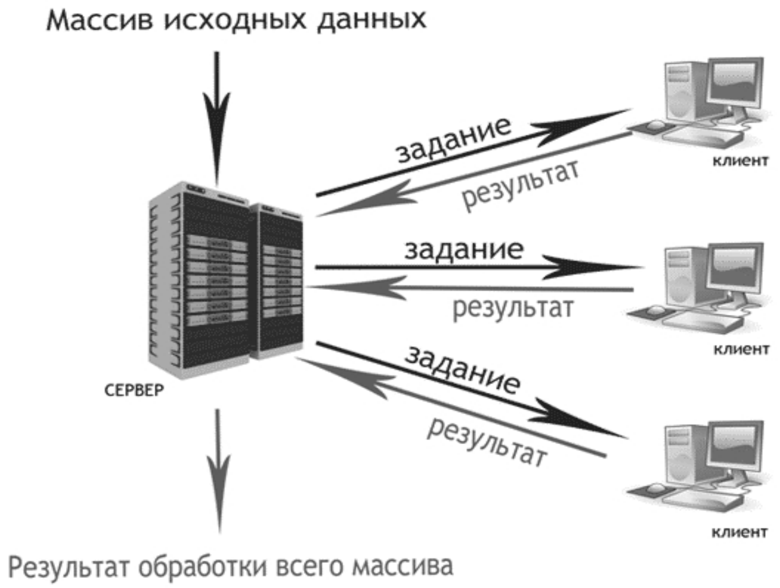


Рисунок 1 – Упрощенная схема распределенной вычислительной сети

С точки зрения системного подхода, распределенная вычислительная сеть представляет собой сложную систему, то есть систему совокупность разнотипных элементов, объединенных разнотипными связями.

Таким образом, в качестве объекта наблюдения в распределенной вычислительной сети может выступать как отдельный компьютер, так и локальная вычислительная сеть, а признаками могут быть различные количественные характеристики, такие, к примеру, как количество запросов в данный период времени к размещенной в распределенной вычислительной сети файлов, число неудачных попыток входа в систему, а также загрузка центрального процессора того или иного отдельного компьютера.

Учитывая, что состояние любого элемента в некоторый момент времени какой-либо системы может быть описано в виде вектора признаков, то состояние системы в целом может описываться матрицей «объект-признак», являющейся одной из двух разновидностей матриц исходных данных в задачах кластеризации [1]. При этом следует также учитывать, что все объекты – элементы распределенной вычислительной сети должны быть описаны в едином признаковом пространстве, а так как значения какого-либо признака для некоторых элементов может варьироваться в интервале даже в отдельно взятый момент времени, то каждый элемент системы должен быть представлен в виде вектора интервалов значений какого-либо признака. Таким образом, при классификации элементов распределенной вычислительной сети с целью обнаружения аномального поведения того или иного элемента, возникает задача обработки интервально-значных данных.

Основные понятия эвристического метода возможностной кластеризации

Эвристический метод возможностной кластеризации состоит в построении так называемого распределения по c нечетких α-кластеров, являющегося частным случаем возможностного разбиения, в общем случае, определяемого условием

$$\upsilon_{li} \geq 0,\ \sum_{l=1}^{c} \upsilon_{li} > 0,\ i = 1,\ldots,n,\ l = 1,\ldots,c, \qquad (1)$$

где $X = \{x_1, ..., x_n\}$ – исследуемая совокупность объектов, на которой определена нечеткая толерантность T, то есть симметричное, рефлексивное, нетранзитивное бинарное нечеткое отношение, с функцией принадлежности $\mu_T(x_i, x_j)$, $i, j = 1, \ldots, n$, так что строки или столбцы этой нечеткой толерантности являются нечеткими множествами $\{A^1, ..., A^n\}$. В таком случае, для некоторого значения α, $\alpha \in (0,1]$, нечеткое множество уровня α, определяемое условием $A^l_{(\alpha)} = \{(x_i, \mu_{A^l}(x_i)) \mid \mu_{A^l}(x_i) \geq \alpha\}$, $l \in \{1, \ldots, n\}$, такое, что $A^l_{(\alpha)} \subseteq A^l$, $A^l \in \{A^1, \ldots, A^n\}$, будет именоваться нечетким α-кластером с функцией принадлежности υ_{li} объекта $x_i \in X$ нечеткому α-кластеру $A^l_{(\alpha)}$, определяемой выражением

$$\upsilon_{li} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A^l_\alpha \\ 0, & иначе \end{cases}, \qquad (2)$$

где $A^l_\alpha = \{x_i \in X \mid \mu_{A^l}(x_i) \geq \alpha\}$ – α-уровень A^l, $l \in \{1, \ldots, n\}$.

Если условие (1) выполняется для всех $A^l_{(\alpha)} \in R^\alpha_z(X)$, где $R^\alpha_z(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}, 2 \leq c \leq n\}$ – семейство c нечетких α-кластеров для некоторого значения α, порожденных заданной на X нечеткой толерантностью T, то это семейство является распределением множества классифицируемых объектов X по c нечетким α-кластерам. Условие (1) в рассматриваемом случае требует, чтобы все объекты совокупности X были распределены по c нечетким α-кластерам $\{A^1_{(\alpha)}, ..., A^c_{(\alpha)}\}$ с положительными значениями типичности υ_{li}, $l = 1, \ldots, c$, $i = 1, \ldots, n$.

Сущность эвристических алгоритмов возможностной кластеризации заключается в нахождении единственного распределения $R^*(X)$ по априори заданному или нет, числу c нечетких α-кластеров. Эвристические алгоритмы возможностной кластеризации условно подразделяются на два типа: реляционные и алгоритмы, основанные на вычислении прототипов кластеров. В первом случае матрицей исходных данных служит матрица нечеткой толерантности T, являющаяся разновидностью матрицы «объект-объект», а во втором – матрица вида «объект-признак». Семейство реляционных эвристических алгоритмов возможностной кластеризации включает:

– D-AFC(c)-алгоритм [3]: построение распределения $R^*(X)$ по априори заданному числу c частично разделенных нечетких α-кластеров;

– D-AFC-PS(c)-алгоритм [4]: модификация D-AFC(c)-алгоритма, использующая аппарат частичного обучения;

– D-PAFC-алгоритм [5]: построение главного распределения $R^*_P(X)$ по априори неизвестному наименьшему числу c полностью разделенных нечетких α-кластеров.

Необходимо указать, что в работе [6] предложен ряд показателей валидности числа c нечетких α-кластеров в искомом распределении $R^*(X)$, предназначенных для использования совместно с D-AFC(c)-алгоритмом.

С другой стороны, семейство эвристических алгоритмов возможностной кластеризации, основанных на вычислении прототипов, включает [7]:

– D-AFC-TC-алгоритм: построение распределения $R^*(X)$ по априори неизвестному числу c полностью разделенных нечетких α-кластеров;

– D-PAFC-TC-алгоритм: построение главного распределения $R^*_P(X)$ по априори неизвестному наименьшему числу c полностью разделенных нечетких α-кластеров;

– D-AFC-TC(α*)-алгоритм: построение распределения $R^*(X)$ по априори неизвестному числу c полностью разделенных нечетких α -кластеров для априори заданного наименьшего порога сходства α^*.

Следует указать, что эвристические возможностные кластер-процедуры, основанные на вычислении прототипов нечетких α -кластеров, используют транзитивное замыкание $\tilde{T}$ нечеткой толерантности T , и не требуют априорного задания числа c полностью разделенных нечетких α -кластеров в искомом распределении $R^*(X)$. Кроме того, реляционные эвристические возможностные кластер-процедуры являются эффективным средством быстрого прототипирования систем нечеткого вывода [9].

Методология обнаружения аномальных наблюдений в интервально-значных данных

В работе [10] предложена методология применения эвристических алгоритмов возможностной кластеризации к обнаружению аномальных наблюдений, в специальной литературе именуемых также «выбросами», в случае, когда данные об исследуемой совокупности представлены векторами интервалов. В основе предложенной в [10] методологии лежит техника построения устойчивой кластерной структуры, детально изложенная в [11]. Однако перед изложением методологии обнаружения «выбросов» в интервально-значных данных представляется целесообразным кратко напомнить основные методы предварительной обработки данных такого типа, рассмотренные в [12].

Пусть $X = \{x_1,...,x_n\}$ – множество объектов, так что каждый объект x_i описывается m_1 числом признаков, и может быть представлен в виде вектора $x_i = (x_i^1,\ldots,x_i^{t_1},\ldots,x_i^{m_1})$, где $\hat{x}_i^{t_1} \in [\hat{x}_i^{t_1(\min)}, \hat{x}_i^{t_1(\max)}]$. Таким образом, интервально-значные данные могут быть представлены в виде матрицы $\hat{X}_{n\times m_1} = [\hat{x}_i^{t_1(t_2)}]$, $i=1,\ldots,n$, $t_1=1,\ldots,m_1$, $t_2 \in \{\min,\max\}$, которая может быть обработана с помощью обобщенной унитаризации

$$x_i^{t_1(t_2)} = \frac{\hat{x}_i^{t_1(t_2)} - \min\limits_{i,t_2} \hat{x}_i^{t_1(t_2)}}{\max\limits_{i,t_2} \hat{x}_i^{t_1(t_2)} - \min\limits_{i,t_2} \hat{x}_i^{t_1(t_2)}}, \tag{3}$$

так что каждый объект x_i, $i=1,\ldots,n$ множества $X = \{x_1,\ldots,x_n\}$ может рассматриваться как интервально-значное нечеткое множество с функцией принадлежности

$$\mu_{x_i}(x^{t_1}) = [\mu_{x_i}(x^{t_1(\min)}), \mu_{x_i}(x^{t_1(\max)})],\ i=1,\ldots,n\ ,\ t_1=1,\ldots,m_1 .$$

Для интервально-значных нечетких множеств рядом авторов были предложены различные расстояния и меры близости – в частности, в работе [13] П. Бурило и Г. Бустинцем было предложено нормализованное евклидово расстояние, определяемое выражением

$$d_I(x_i,x_j) = \sqrt{\frac{1}{2m_1}\sum_{t_1=1}^{m_1}\left(\left(\mu_{x_i}(x^{t_1(\min)}) - \mu_{x_j}(x^{t_1(\min)})\right)^2 + \left(\mu_{x_i}(x^{t_1(\max)}) - \mu_{x_j}(x^{t_1(\max)})\right)^2\right)}, \tag{4}$$

для всех $i,j=1,\ldots,n$. С другой стороны, обобщение относительного евклидова расстояния, предложенного в [14], для случая интервально-значных нечетких множеств примет вид

$$e_I(x_i, x_j) = \sqrt{\frac{1}{m_1}\sum_{t_1=1}^{m_1}\left(\frac{1}{2^2}\sum_{t_2\in\{\min,\max\}}\left(\mu_{x_i}(x^{t_1(t_2)}) - \mu_{x_j}(x^{t_1(t_2)})\right)^2\right)}\,, \tag{5}$$

также для всех $i, j = 1,\ldots,n$. Как отмечалось в [12], построенное с помощью формулы (16) нечеткое отношение несходства сохраняет только свойство симметричности, так что обобщение относительного евклидова расстояния (5) представляет собой меру различия.

В результате применения расстояния (4) или меры различия (5) к матрице нормированных в соответствии с формулой (3) интервально-значных данных, получается матрица нечеткого отношения несходства $I_{n\times n} = [\mu_I(x_i, x_j)]$, применение к которой, в свою очередь, операции дополнения дает в результате матрицу нечеткой толерантности $T_{n\times n} = [\mu_T(x_i, x_j)]$, являющуюся матрицей исходных данных для реляционных алгоритмов эвристического метода возможностной кластеризации.

Кроме того, П. Гжегожевским в [15] было предложено расстояние между интервально-значными нечеткими множествами, основанное на метрике Хаусдорфа, а Х. Юу и Х. Юаном в [16] была определена мера близости интервально-значных нечетких множеств, которые подробно рассмотрены в [12].

С содержательной точки зрения, аномальное наблюдение представляет собой отдельный элемент, находящийся на значительном удалении от кластеров и других отдельных элементов [8]. Таким образом, предложенная в [10] методология обнаружения аномальных наблюдений в интервально-значных данных может быть описана в виде следующей последовательности шагов:

1. В соответствии с предложенной в [11] методологией построения устойчивой кластерной структуры, распределение $R^*(X)$ по априори неизвестному числу c нечетких α-кластеров;

2. Для каждого нечеткого α-кластера $A^l_{(\alpha)} \in R^*(X)$ вычисляется мощность его носителя, $card(A^l_\alpha)$;

3. Производится проверка условия: **если** $card(A^l_\alpha) = 1$ **то** этот нечеткий α-кластер, носитель которого содержит единственный элемент $x_i \in X$, является аномальным наблюдением.

Следует отметить, что если под аномальным наблюдением понимать немногочисленную, по сравнению с остальными, группу объектов, чему соответствует нечеткий α-кластер, мощность носителя которого превышает 1, то предложенная методология может быть тривиально обобщена на указанный случай.

Иллюстративный пример

В качестве иллюстрации применения предложенной методологии к решению задачи обнаружения аномального поведения элементов распределенной вычислительной сети может послужить рассмотренный в [10] пример двумерных интервально-значных данных, приведенных на рис. 2.

Очевидно, что в исследуемой совокупности $X = \{x_1,\ldots,x_{16}\}$ группы объектов $\{x_1,\ldots,x_6\}$ и $\{x_7,\ldots,x_{12}\}$ образуют кластеры, а объекты x_{13}, x_{14}, x_{15} и x_{16} являются «выбросами».

Результат применения изложенной методологии к задаче обнаружения «выбросов» представлен на рис. 3.

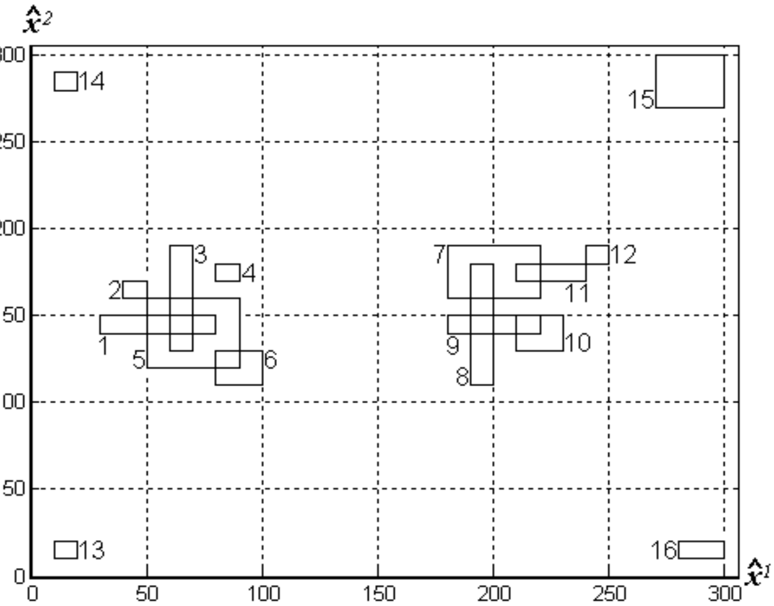
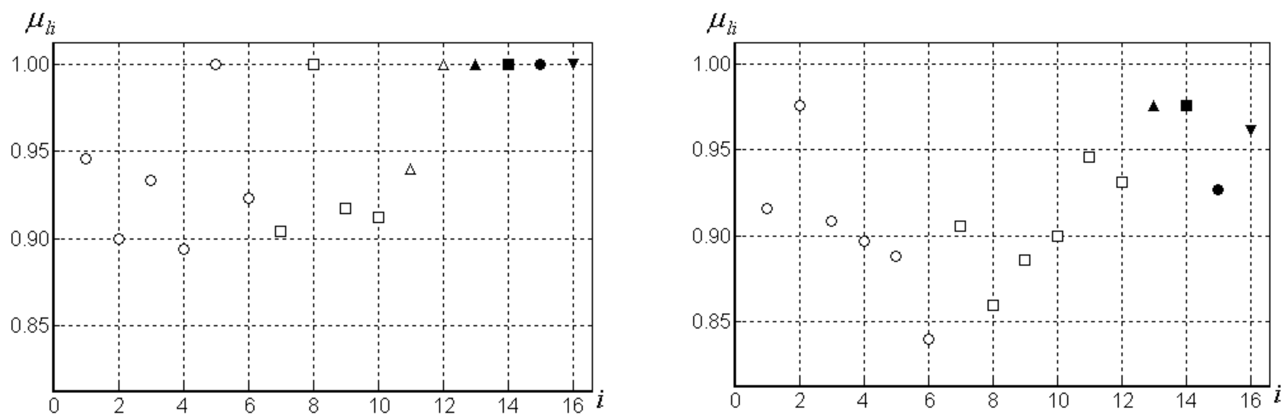


Рисунок 2 – Интервально-значные данные для вычислительного эксперимента



а) б)

Рисунок 3 – Значения принадлежностей объектов классам полученного распределения: при использовании расстояния (4) (а), при использовании меры различия (5) (б)

При использовании в описанной методологии расстояния (4), группа объектов $\{x_7,...,x_{12}\}$ разделилась на две подгруппы: $\{x_7,...,x_{10}\}$ и $\{x_{11},x_{12}\}$. С другой стороны, использование меры различия (5) выделяет две группы $\{x_1,...,x_6\}$ и $\{x_7,...,x_{12}\}$, а объекты x_{13}, x_{14}, x_{15} и x_{16} выделяются в отдельные классы в обоих случаях и могут интепретироваться как аномальные наблюдения. Следует также указать, что в случае использования меры различия (5) все нечеткие α -кластеры представляют собой субнормальные нечеткие множества. Результаты использования расстояния П. Гжегожевского [15] и меры близости Х. Юу и Х. Юана [16] приведены в [10] и сходны с приведенными на рис. 3. результатами.

Выводы

В работе предложен подход к обнаружению аномального поведения в распределенных вычислительных сетях, основанный на представлении данных о состоянии элементов системы в виде векторов интервалов и методологии обнаружения «выбросов» в интервально-значных данных. Результаты вычислительного эксперимента демонстрируют эффективность предложенного подхода.

Предложенный подход также может быть обобщен на случай представления данных о состоянии элементов системы в виде векторов нечетких чисел [17], что представляет интерес как с теоретической, так и с практической точек зрения, и является направлением перспективных исследований.

Литература

1. Мандель И.Д. Кластерный анализ / И.Д. Мандель. – М. : Финансы и статистика, 1988. – 176 с.
2. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition / F. Höppner [et al].– Chichester : Wiley Intersciences, 1999.– 289 p.
3. Viattchenin D.A. A new heuristic algorithm of fuzzy clustering / D.A. Viattchenin // Control & Cybernetics. – 2004. – Vol. 33. – P. 323-340.
4. Viattchenin D.A. A direct algorithm of possibilistic clustering with partial supervision / D.A. Viattchenin // Journal of Automation, Mobile Robotics and Intelligent Systems. – 2007. – Vol. 1, № 3. – P. 29-38.
5. Viattchenin D.A. An algorithm for detecting the principal allotment among fuzzy clusters and its application as a technique of reduction of analyzed features space dimensionality / D.A. Viattchenin // Journal of Information and Organizational Sciences. – 2009. – Vol. 33. – P. 205-217.
6. Viattchenin D.A. Validity measures for heuristic possibilistic clustering / D.A. Viattchenin // Information Technology and Control. – 2010. – Vol. 39, № 4. – P. 321-332.
7. Вятченин Д.А. Прямые алгоритмы нечеткой кластеризации, основанные на операции транзитивного замыкания и их применение к обнаружению аномальных наблюдений / Д.А. Вятченин // Искусственный интеллект. – 2007. – № 3. – С. 205-216.
8. Методология применения эвристических алгоритмов возможностной кластеризации в системах обнаружения атак / В.В. Анищенко, Д.А. Вятченин, А.В. Доморацкий, Р. Тати // Безопасность информационных технологий. – 2012. – № 1. – С. 19-21.
9. Анищенко В.В. Обучение консеквентов нечетких правил, построенных на основе эвристической возможностной кластеризации / В.В. Анищенко, Д.А. Вятченин, А.В. Доморацкий // Информатика. – 2011. – № 3. – С. 98-111.
10. Viattchenin D.A. Detecting outliers in interval-valued data using heuristic possibilistic clustering / D.A. Viattchenin // Journal of Computer Science and Control Systems. – 2012. – Vol.5, № 2. (to appear)
11. Viattchenin D.A. Constructing stable clustering structure for uncertain data set / D.A. Viattchenin // Acta Electrotechnica et Informatica. – 2011. – Vol. 11, № 3. – P. 42-50.
12. Вятченин Д.А. Построение нечеткого с-разбиения в случае неустойчивой кластерной структуры множества объектов / Д.А. Вятченин, А.В. Доморацкий // Искусственный интеллект. –2011. – № 3. – С. 479-489.
13. Burillo P. Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets / P. Burillo, H. Bustince // Fuzzy Sets and Systems. – 1996. – Vol. 78. – P. 305-316.
14. Viattchenin D.A. An outline for a heuristic approach to possibilistic clustering of the three-way data / D.A. Viattchenin // Journal of Uncertain Systems. – 2009. – Vol. 3. – P. 64-80.
15. Grzegorzewski P. Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on Hausdorff metric // Fuzzy Sets and Systems. – 2004. – Vol. 148. – P. 319-328.
16. Ju H. Similarity measures on interval-valued fuzzy sets and application to pattern recognition / H. Ju, X. Yan // Fuzzy Information and Engineering / Ed. by D.Y. Cao. – Berlin: Springer-Verlag, 2007. – P. 875–883.
17. Фисенко В.К. Формализация нечеткого описания признаков объектов информатизации в задаче категорирования по требованиям информационной безопасности / Фисенко В.К., Д.А. Вятченин // Искусственный интеллект. Интеллектуальные системы ИИ-2011 : материалы Международной научно-технической конференции (пос. Кацивели, АР Крым, 19 – 23 сентября 2011 года). – Донецк : Наука і освіта, 2011. – С. 302-306.

Literatura

1. Mandel I.D. Cluster Analysis. Moscow: Finansy i Statistika. 1988. 176 p. (in Russian)
2. Höppner F. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. Chichester: Wiley Intersciences. 1999. 289 p.
3. Viattchenin D.A. Control & Cybernetics. 2004. Vol. 33. P. 323-340.
4. Viattchenin D.A. Journal of Automation, Mobile Robotics and Intelligent Systems. 2007. Vol. 1. № 3. P. 29-38.

5. Viattchenin D.A. Journal of Information and Organizational Sciences. 2009. Vol. 33. P. 205-217.
6. Viattchenin D.A. Information Technology and Control. 2010. Vol. 39, № 4. P. 321-332.
7. Viattchenin D.A. Iskusstvennyj intellect. 2007. №.3. S. 205-216. (in Russian)
8. Anishchanka U.V. Bezopasnost’ informatsionnyh technologij. 2012. № 1. S. 19-21. (in Russian)
9. Anishchanka U.V. Informatics. 2011. №.3. S. 98-111. (in Russian)
10. Viattchenin D.A. Journal of Computer Science and Control Systems. 2012. Vol.5, № 2. (to appear)
11. Viattchenin D.A. Acta Electrotechnica et Informatica. 2011. Vol.11, № 3. P. 42-50.
12. Viattchenin D.A. Iskusstvennyj intellect. 2011. №.3. S. 479-489. (in Russian)
13. Burillo P. Fuzzy Sets and Systems. 1996. Vol. 78. P. 305-316.
14. Viattchenin D.A. Journal of Uncertain Systems. 2009. Vol. 3. P. 64-80.
15. Grzegorzewski P. Fuzzy Sets and Systems. 2004. Vol. 148. P. 319-328.
16. Ju H. Fuzzy Information and Engineering. Berlin: Springer-Verlag. 2007. P. 875–883.
17. Fisenko V.K. Iskusstvennyj intellect. Intellectualnyje systemy. Donetsk: Nauka i osvita. 2011. S. 302-306. (in Russian)

RESUME

U.V. Anishchenko, D.A. Viattchenin, A.V. Damaratski, V.K. Fisenko

Detecting Anomalous Behavior via Discovering Attacks Systems for Interval-Valued Data

The paper deals with the problem of detecting anomalous user behavior in a distributed computational network. The data should be represented as a table where each cell of this table contains an interval of values.

The method of detecting outliers in interval-valued data was applied for the problem solving. The method is based on constructing stable clustering structure using a heuristic method of possibilistic clustering. Results of numerical experiment seem to be satisfactory. So, the proposed approach is effective tool for detecting anomalous user behavior in computational networks.

Статья поступила в редакцию 05.06.2012.

УДК 004,7.056

В.В. Анищенко, Д.А. Вятченин, А.В. Доморацкий, В.К. Фисенко
Объединенный институт проблем информатики НАН Беларуси, г. Минск
Беларусь, 220012, г. Минск, ул. Сурганова, 6

Выявление аномального поведения системами обнаружения атак при интервально-значном представлении данных

V.V. Anishchenko, D.A. Viattchenin, A.V. Damaratski, V.K. Fisenko
United Institute of Informatics Problems
NAS of Belarus, c. Minsk
Belarus, 220012, c. Minsk, Surganov st., 6

Detecting Anomalous Behavior via Discovering Attacks Systems for Interval-Valued Data

В.В. Анищенко, Д.А. Вятченін, А.В. Доморацький, В.К. Фісенко
Об’єднаний інститут проблем інформатики НАН Білорусі, м. Мінськ
Білорусь, 220012, м. Мінськ, вул. Сурганова, 6

Вияв аномальної поведінки системами виявлення атак при інтервально-значному зображенні даних

В статье рассмотрен метод обнаружения аномального поведения пользователей распределенной компьютерной сети при интервально-значном представлении данных, основанный на построении устойчивой кластерной структуры с помощью эвристического метода возможностной кластеризации. Предложенный метод иллюстрируется результатами вычислительного эксперимента.
Ключевые слова: кластеризация, аномальные наблюдения, неопределенные данные.

A method of detecting anomalous user behavior in a distributed computational network for a case of interval-valued data is considered in the article. The method is based on constructing stable clustering structure using a heuristic method of possibilistic clustering. The proposed method is illustrated by the results of numerical experiment.
Key words: clustering, anomalous observations, uncertain data.

У статті розглянуто метод виявлення аномальної поведінки користувачів розподіленої комп’ютерної мережі при інтервально-значному зображенні даних, що заснований на побудові стійкої кластерної структури за допомогою евристичного методу можливісної кластеризації. Запропонований метод ілюструється результатами обчислювального експерименту.
Ключові слова: кластеризація, аномальні спостереження, невизначені дані.

Введение

Системы обнаружения атак на объекты информатизации давно применяются как один из необходимых рубежей защиты информационных систем. Системы обнаружения атак представляют собой программные или аппаратно-программные решения, которые автоматизируют процесс контроля событий, протекающих в компьютерной системе или сети, а также самостоятельно анализируют эти события в поисках признаков проблем

безопасности. Так как количество различных типов и способов организации несанкционированных проникновений в чужие сети значительно увеличилось за последние годы, системы обнаружения атак стали необходимым компонентом инфраструктуры безопасности большинства организаций.

Системы обнаружения атак условно делятся на два типа: системы обнаружения злоумышленного поведения и системы обнаружения аномального поведения. Системы обнаружения злоумышленного поведения основаны на информации о признаках, характеризующих поведение злоумышленника, тогда как работа систем обнаружения аномального поведения основана на информации о некоторых признаках, характеризующих допустимое поведение объекта наблюдения, где под допустимым поведением понимаются действия, выполняемые объектом и не противоречащие политике безопасности. Главным достоинством систем обнаружения аномального поведения является возможность генерирования системами указанного типа информации, которая может быть использована в системах обнаружения злоумышленного поведения, что, в свою очередь, открывает возможности создания гибридных систем обнаружения атак.

Наиболее распространенным видом реализации технологии обнаружения аномального поведения является применение различных статистических методов, в том числе кластерного анализа [1]. Следует также указать, что в работе [1] особо отмечается высокая эффективность методов кластеризации в задачах обнаружения аномальных наблюдений в исследуемой совокупности объектов. В последние годы особый интерес у исследователей вызывают методы нечеткой и возможностной кластеризации [2], отличительной чертой которых является не просто указание принадлежности того или иного объекта к определенному кластеру, но и степень, с которой данный объект принадлежит тому или иному таксону. Необходимо отметить, что подавляющее большинство алгоритмов нечеткой и возможностной кластеризации являются представителями так называемого оптимизационного направления. С другой стороны, в работах [3-7] предложен так называемый эвристический подход к решению задачи возможностной кластеризации, отличающийся от оптимизационного подхода устойчивостью результатов классификации.

В работе [8] предложена методология применения эвристических алгоритмов возможностной кластеризации при разработке систем обнаружения аномального поведения, которая основывается на выработке признакового пространства, описывающего нормальное поведение объектов информатизации, с последующим сбором статистической информации и обнаружением объектов, поведение которых отличается от допустимого. Недостатком предложенного в [8] подхода является описание поведения объектов информатизации в виде вектора некоторых количественных признаков, что является приемлемым для компьютерных систем, насчитывающих сравнительно небольшое число однотипных элементов, примером которых являются локальные вычислительные сети.

Целью данной работы является модификация предложенной в [8] методологии для случая распределенных вычислительных систем, поведение элементов которых может описываться векторами интервалов.

Представление данных о поведении элементов распределенной вычислительной сети

Распределенные вычислительные системы представляют собой совокупность значительно удаленных друг от друга отдельных ЭВМ и локальных сетей, представляющих собой вычислительные узлы. Распределенные вычислительные системы используются

для решения как наборов независимых задач, так и единой сложной задачи. На рис. 1 приведена упрощенная схема организации распределенной вычислительной сети.

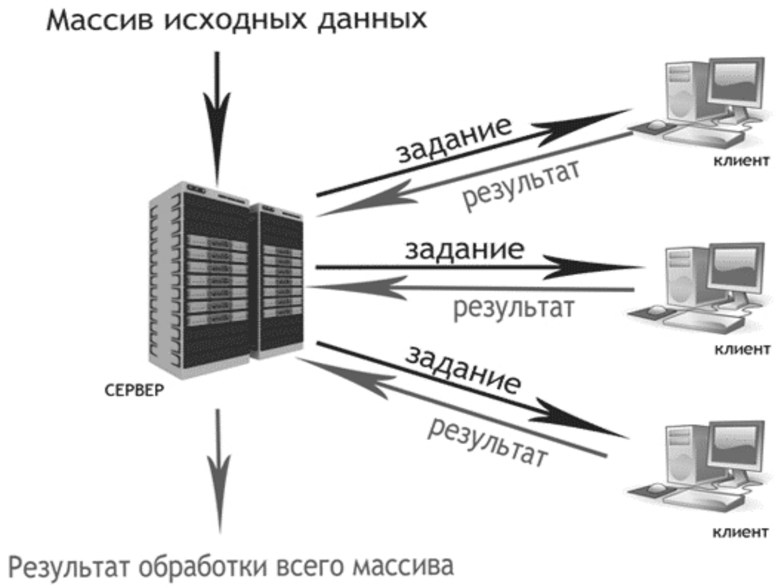


Рисунок 1 – Упрощенная схема распределенной вычислительной сети

С точки зрения системного подхода, распределенная вычислительная сеть представляет собой сложную систему, то есть систему совокупность разнотипных элементов, объединенных разнотипными связями.

Таким образом, в качестве объекта наблюдения в распределенной вычислительной сети может выступать как отдельный компьютер, так и локальная вычислительная сеть, а признаками могут быть различные количественные характеристики, такие, к примеру, как количество запросов в данный период времени к размещенной в распределенной вычислительной сети файлов, число неудачных попыток входа в систему, а также загрузка центрального процессора того или иного отдельного компьютера.

Учитывая, что состояние любого элемента в некоторый момент времени какой-либо системы может быть описано в виде вектора признаков, то состояние системы в целом может описываться матрицей «объект-признак», являющейся одной из двух разновидностей матриц исходных данных в задачах кластеризации [1]. При этом следует также учитывать, что все объекты – элементы распределенной вычислительной сети должны быть описаны в едином признаковом пространстве, а так как значения какого-либо признака для некоторых элементов может варьироваться в интервале даже в отдельно взятый момент времени, то каждый элемент системы должен быть представлен в виде вектора интервалов значений какого-либо признака. Таким образом, при классификации элементов распределенной вычислительной сети с целью обнаружения аномального поведения того или иного элемента, возникает задача обработки интервально-значных данных.

Основные понятия эвристического метода возможностной кластеризации

Эвристический метод возможностной кластеризации состоит в построении так называемого распределения по c нечетких α-кластеров, являющегося частным случаем возможностного разбиения, в общем случае, определяемого условием

$$\upsilon_{li} \geq 0,\ \sum_{l=1}^{c} \upsilon_{li} > 0,\ i = 1, \ldots, n,\ l = 1, \ldots, c, \tag{1}$$

где $X = \{x_1, ..., x_n\}$ – исследуемая совокупность объектов, на которой определена нечеткая толерантность T, то есть симметричное, рефлексивное, нетранзитивное бинарное нечеткое отношение, с функцией принадлежности $\mu_T(x_i, x_j)$, $i, j = 1, \ldots, n$, так что строки или столбцы этой нечеткой толерантности являются нечеткими множествами $\{A^1, ..., A^n\}$. В таком случае, для некоторого значения α, $\alpha \in (0,1]$, нечеткое множество уровня α, определяемое условием $A^l_{(\alpha)} = \{(x_i, \mu_{A^l}(x_i)) \mid \mu_{A^l}(x_i) \geq \alpha\}$, $l \in \{1, \ldots, n\}$, такое, что $A^l_{(\alpha)} \subseteq A^l$, $A^l \in \{A^1, \ldots, A^n\}$, будет именоваться нечетким α-кластером с функцией принадлежности υ_{li} объекта $x_i \in X$ нечеткому α-кластеру $A^l_{(\alpha)}$, определяемой выражением

$$\upsilon_{li} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A^l_\alpha \\ 0, & \textit{иначе} \end{cases}, \tag{2}$$

где $A^l_\alpha = \{x_i \in X \mid \mu_{A^l}(x_i) \geq \alpha\}$ – α-уровень A^l, $l \in \{1, \ldots, n\}$.

Если условие (1) выполняется для всех $A^l_{(\alpha)} \in R^\alpha_z(X)$, где $R^\alpha_z(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}, 2 \leq c \leq n\}$ – семейство c нечетких α-кластеров для некоторого значения α, порожденных заданной на X нечеткой толерантностью T, то это семейство является распределением множества классифицируемых объектов X по c нечетким α-кластерам. Условие (1) в рассматриваемом случае требует, чтобы все объекты совокупности X были распределены по c нечетким α-кластерам $\{A^1_{(\alpha)}, ..., A^c_{(\alpha)}\}$ с положительными значениями типичности υ_{li}, $l = 1, \ldots, c$, $i = 1, \ldots, n$.

Сущность эвристических алгоритмов возможностной кластеризации заключается в нахождении единственного распределения $R^*(X)$ по априори заданному или нет, числу c нечетких α-кластеров. Эвристические алгоритмы возможностной кластеризации условно подразделяются на два типа: реляционные и алгоритмы, основанные на вычислении прототипов кластеров. В первом случае матрицей исходных данных служит матрица нечеткой толерантности T, являющаяся разновидностью матрицы «объект-объект», а во втором – матрица вида «объект-признак». Семейство реляционных эвристических алгоритмов возможностной кластеризации включает:

– D-AFC(c)-алгоритм [3]: построение распределения $R^*(X)$ по априори заданному числу c частично разделенных нечетких α-кластеров;

– D-AFC-PS(c)-алгоритм [4]: модификация D-AFC(c)-алгоритма, использующая аппарат частичного обучения;

– D-PAFC-алгоритм [5]: построение главного распределения $R^*_P(X)$ по априори неизвестному наименьшему числу c полностью разделенных нечетких α-кластеров.

Необходимо указать, что в работе [6] предложен ряд показателей валидности числа c нечетких α-кластеров в искомом распределении $R^*(X)$, предназначенных для использования совместно с D-AFC(c)-алгоритмом.

С другой стороны, семейство эвристических алгоритмов возможностной кластеризации, основанных на вычислении прототипов, включает [7]:

– D-AFC-TC-алгоритм: построение распределения $R^*(X)$ по априори неизвестному числу c полностью разделенных нечетких α-кластеров;

– D-PAFC-TC-алгоритм: построение главного распределения $R^*_P(X)$ по априори неизвестному наименьшему числу c полностью разделенных нечетких α-кластеров;

– D-AFC-TC(α*)-алгоритм: построение распределения $R^*(X)$ по априори неизвестному числу c полностью разделенных нечетких α -кластеров для априори заданного наименьшего порога сходства α^*.

Следует указать, что эвристические возможностные кластер-процедуры, основанные на вычислении прототипов нечетких α -кластеров, используют транзитивное замыкание $\tilde{T}$ нечеткой толерантности T, и не требуют априорного задания числа c полностью разделенных нечетких α -кластеров в искомом распределении $R^*(X)$. Кроме того, реляционные эвристические возможностные кластер-процедуры являются эффективным средством быстрого прототипирования систем нечеткого вывода [9].

Методология обнаружения аномальных наблюдений в интервально-значных данных

В работе [10] предложена методология применения эвристических алгоритмов возможностной кластеризации к обнаружению аномальных наблюдений, в специальной литературе именуемых также «выбросами», в случае, когда данные об исследуемой совокупности представлены векторами интервалов. В основе предложенной в [10] методологии лежит техника построения устойчивой кластерной структуры, детально изложенная в [11]. Однако перед изложением методологии обнаружения «выбросов» в интервально-значных данных представляется целесообразным кратко напомнить основные методы предварительной обработки данных такого типа, рассмотренные в [12].

Пусть $X = \{x_1,..., x_n\}$ – множество объектов, так что каждый объект x_i описывается m_1 числом признаков, и может быть представлен в виде вектора $x_i = (x_i^1,\ldots,x_i^{t_1},\ldots,x_i^{m_1})$, где $\hat{x}_i^{t_1} \in [\hat{x}_i^{t_1(\min)}, \hat{x}_i^{t_1(\max)}]$. Таким образом, интервально-значные данные могут быть представлены в виде матрицы $\hat{X}_{n\times m_1} = [\hat{x}_i^{t_1(t_2)}]$, $i = 1,\ldots,n$, $t_1 = 1,\ldots,m_1$, $t_2 \in \{\min, \max\}$, которая может быть обработана с помощью обобщенной унитаризации

$$x_i^{t_1(t_2)} = \frac{\hat{x}_i^{t_1(t_2)} - \min\limits_{i,t_2} \hat{x}_i^{t_1(t_2)}}{\max\limits_{i,t_2} \hat{x}_i^{t_1(t_2)} - \min\limits_{i,t_2} \hat{x}_i^{t_1(t_2)}}, \tag{3}$$

так что каждый объект x_i, $i = 1,\ldots,n$ множества $X = \{x_1,\ldots,x_n\}$ может рассматриваться как интервально-значное нечеткое множество с функцией принадлежности

$$\mu_{x_i}(x^{t_1}) = [\mu_{x_i}(x^{t_1(\min)}), \mu_{x_i}(x^{t_1(\max)})],\ i = 1,\ldots,n\ ,\ t_1 = 1,\ldots,m_1 .$$

Для интервально-значных нечетких множеств рядом авторов были предложены различные расстояния и меры близости – в частности, в работе [13] П. Бурило и Г. Бустинцем было предложено нормализованное евклидово расстояние, определяемое выражением

$$d_I(x_i, x_j) = \sqrt{\frac{1}{2m_1}\sum_{t_1=1}^{m_1}\left(\left(\mu_{x_i}(x^{t_1(\min)}) - \mu_{x_j}(x^{t_1(\min)})\right)^2 + \left(\mu_{x_i}(x^{t_1(\max)}) - \mu_{x_j}(x^{t_1(\max)})\right)^2\right)}, \tag{4}$$

для всех $i, j = 1,\ldots,n$. С другой стороны, обобщение относительного евклидова расстояния, предложенного в [14], для случая интервально-значных нечетких множеств примет вид

$$e_I(x_i, x_j) = \sqrt{\frac{1}{m_1}\sum_{t_1=1}^{m_1}\left(\frac{1}{2^2}\sum_{t_2\in\{\min,\max\}}\left(\mu_{x_i}(x^{t_1(t_2)}) - \mu_{x_j}(x^{t_1(t_2)})\right)^2\right)}, \qquad (5)$$

также для всех $i, j = 1, \ldots, n$. Как отмечалось в [12], построенное с помощью формулы (16) нечеткое отношение несходства сохраняет только свойство симметричности, так что обобщение относительного евклидова расстояния (5) представляет собой меру различия.

В результате применения расстояния (4) или меры различия (5) к матрице нормированных в соответствии с формулой (3) интервально-значных данных, получается матрица нечеткого отношения несходства $I_{n\times n} = [\mu_I(x_i, x_j)]$, применение к которой, в свою очередь, операции дополнения дает в результате матрицу нечеткой толерантности $T_{n\times n} = [\mu_T(x_i, x_j)]$, являющуюся матрицей исходных данных для реляционных алгоритмов эвристического метода возможностной кластеризации.

Кроме того, П. Гжегожевским в [15] было предложено расстояние между интервально-значными нечеткими множествами, основанное на метрике Хаусдорфа, а Х. Юу и Х. Юаном в [16] была определена мера близости интервально-значных нечетких множеств, которые подробно рассмотрены в [12].

С содержательной точки зрения, аномальное наблюдение представляет собой отдельный элемент, находящийся на значительном удалении от кластеров и других отдельных элементов [8]. Таким образом, предложенная в [10] методология обнаружения аномальных наблюдений в интервально-значных данных может быть описана в виде следующей последовательности шагов:

1. В соответствии с предложенной в [11] методологией построения устойчивой кластерной структуры, распределение $R^*(X)$ по априори неизвестному числу c нечетких α-кластеров;

2. Для каждого нечеткого α-кластера $A^l_{(\alpha)} \in R^*(X)$ вычисляется мощность его носителя, $card(A^l_\alpha)$;

3. Производится проверка условия: **если** $card(A^l_\alpha) = 1$ **то** этот нечеткий α-кластер, носитель которого содержит единственный элемент $x_i \in X$, является аномальным наблюдением.

Следует отметить, что если под аномальным наблюдением понимать немногочисленную, по сравнению с остальными, группу объектов, чему соответствует нечеткий α-кластер, мощность носителя которого превышает 1, то предложенная методология может быть тривиально обобщена на указанный случай.

Иллюстративный пример

В качестве иллюстрации применения предложенной методологии к решению задачи обнаружения аномального поведения элементов распределенной вычислительной сети может послужить рассмотренный в [10] пример двумерных интервально-значных данных, приведенных на рис. 2.

Очевидно, что в исследуемой совокупности $X = \{x_1,...,x_{16}\}$ группы объектов $\{x_1,...,x_6\}$ и $\{x_7,...,x_{12}\}$ образуют кластеры, а объекты x_{13}, x_{14}, x_{15} и x_{16} являются «выбросами».

Результат применения изложенной методологии к задаче обнаружения «выбросов» представлен на рис. 3.

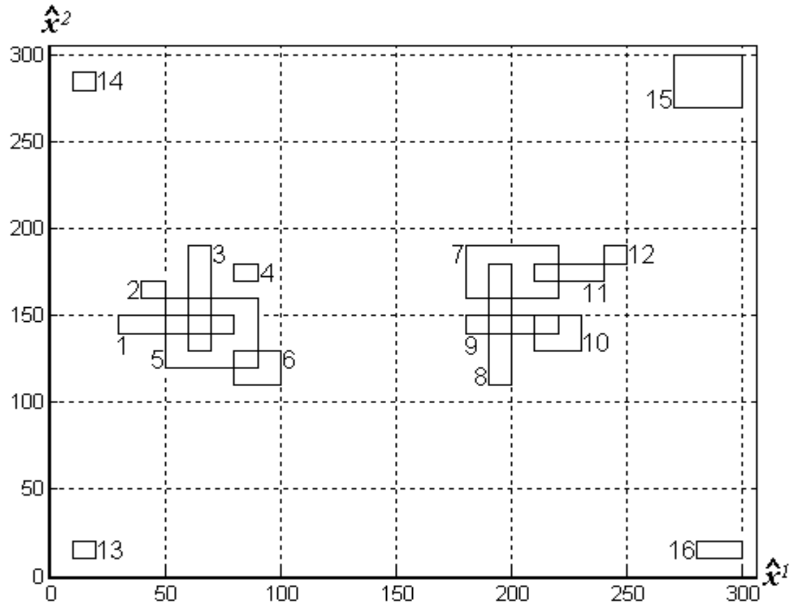
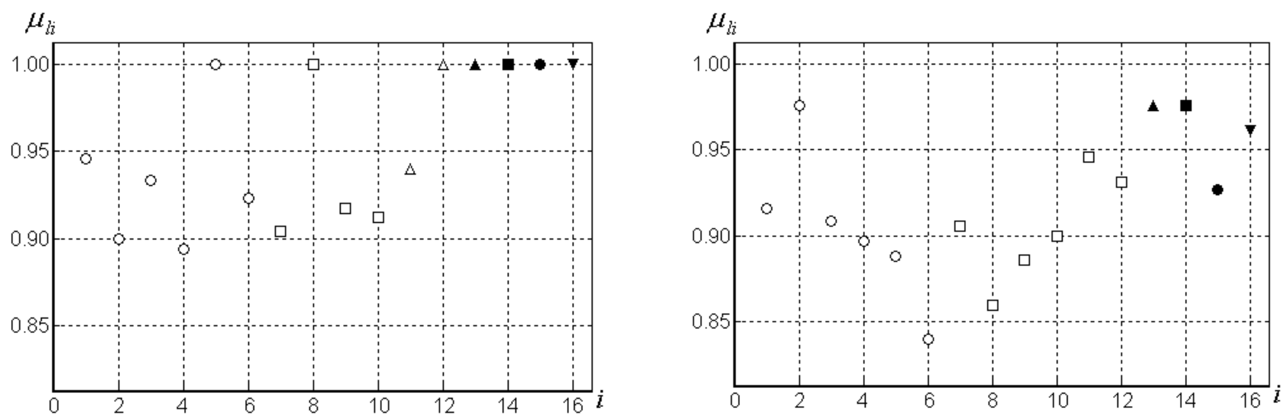


Рисунок 2 – Интервально-значные данные для вычислительного эксперимента



а) б)

Рисунок 3 – Значения принадлежностей объектов классам полученного распределения: при использовании расстояния (4) (а), при использовании меры различия (5) (б)

При использовании в описанной методологии расстояния (4), группа объектов $\{x_7,...,x_{12}\}$ разделилась на две подгруппы: $\{x_7,...,x_{10}\}$ и $\{x_{11},x_{12}\}$. С другой стороны, использование меры различия (5) выделяет две группы $\{x_1,...,x_6\}$ и $\{x_7,...,x_{12}\}$, а объекты x_{13}, x_{14}, x_{15} и x_{16} выделяются в отдельные классы в обоих случаях и могут интепретироваться как аномальные наблюдения. Следует также указать, что в случае использования меры различия (5) все нечеткие α-кластеры представляют собой субнормальные нечеткие множества. Результаты использования расстояния П. Гжегожевского [15] и меры близости Х. Юу и Х. Юана [16] приведены в [10] и сходны с приведенными на рис. 3. результатами.

Выводы

В работе предложен подход к обнаружению аномального поведения в распределенных вычислительных сетях, основанный на представлении данных о состоянии элементов системы в виде векторов интервалов и методологии обнаружения «выбросов» в интервально-значных данных. Результаты вычислительного эксперимента демонстрируют эффективность предложенного подхода.

Предложенный подход также может быть обобщен на случай представления данных о состоянии элементов системы в виде векторов нечетких чисел [17], что представляет интерес как с теоретической, так и с практической точек зрения, и является направлением перспективных исследований.

Литература

1. Мандель И.Д. Кластерный анализ / И.Д. Мандель. – М. : Финансы и статистика, 1988. – 176 с.
2. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition / F. Höppner [et al].– Chichester : Wiley Intersciences, 1999.– 289 p.
3. Viattchenin D.A. A new heuristic algorithm of fuzzy clustering / D.A. Viattchenin // Control & Cybernetics. – 2004. – Vol. 33. – P. 323-340.
4. Viattchenin D.A. A direct algorithm of possibilistic clustering with partial supervision / D.A. Viattchenin // Journal of Automation, Mobile Robotics and Intelligent Systems. – 2007. – Vol. 1, № 3. – P. 29-38.
5. Viattchenin D.A. An algorithm for detecting the principal allotment among fuzzy clusters and its application as a technique of reduction of analyzed features space dimensionality / D.A. Viattchenin // Journal of Information and Organizational Sciences. – 2009. – Vol. 33. – P. 205-217.
6. Viattchenin D.A. Validity measures for heuristic possibilistic clustering / D.A. Viattchenin // Information Technology and Control. – 2010. – Vol. 39, № 4. – P. 321-332.
7. Вятченин Д.А. Прямые алгоритмы нечеткой кластеризации, основанные на операции транзитивного замыкания и их применение к обнаружению аномальных наблюдений / Д.А. Вятченин // Искусственный интеллект. – 2007. – № 3. – С. 205-216.
8. Методология применения эвристических алгоритмов возможностной кластеризации в системах обнаружения атак / В.В. Анищенко, Д.А. Вятченин, А.В. Доморацкий, Р. Тати // Безопасность информационных технологий. – 2012. – № 1. – С. 19-21.
9. Анищенко В.В. Обучение консеквентов нечетких правил, построенных на основе эвристической возможностной кластеризации / В.В. Анищенко, Д.А. Вятченин, А.В. Доморацкий // Информатика. – 2011. – № 3. – С. 98-111.
10. Viattchenin D.A. Detecting outliers in interval-valued data using heuristic possibilistic clustering / D.A. Viattchenin // Journal of Computer Science and Control Systems. – 2012. – Vol.5, № 2. (to appear)
11. Viattchenin D.A. Constructing stable clustering structure for uncertain data set / D.A. Viattchenin // Acta Electrotechnica et Informatica. – 2011. – Vol. 11, № 3. – P. 42-50.
12. Вятченин Д.А. Построение нечеткого с-разбиения в случае неустойчивой кластерной структуры множества объектов / Д.А. Вятченин, А.В. Доморацкий // Искусственный интеллект. –2011. – № 3. – С. 479-489.
13. Burillo P. Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy sets / P. Burillo, H. Bustince // Fuzzy Sets and Systems. – 1996. – Vol. 78. – P. 305-316.
14. Viattchenin D.A. An outline for a heuristic approach to possibilistic clustering of the three-way data / D.A. Viattchenin // Journal of Uncertain Systems. – 2009. – Vol. 3. – P. 64-80.
15. Grzegorzewski P. Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on Hausdorff metric // Fuzzy Sets and Systems. – 2004. – Vol. 148. – P. 319-328.
16. Ju H. Similarity measures on interval-valued fuzzy sets and application to pattern recognition / H. Ju, X. Yan // Fuzzy Information and Engineering / Ed. by D.Y. Cao. – Berlin: Springer-Verlag, 2007. – P. 875–883.
17. Фисенко В.К. Формализация нечеткого описания признаков объектов информатизации в задаче категорирования по требованиям информационной безопасности / Фисенко В.К., Д.А. Вятченин // Искусственный интеллект. Интеллектуальные системы ИИ-2011 : материалы Международной научно-технической конференции (пос. Кацивели, АР Крым, 19 – 23 сентября 2011 года). – Донецк : Наука і освіта, 2011. – С. 302-306.

Literatura

1. Mandel I.D. Cluster Analysis. Moscow: Finansy i Statistika. 1988. 176 p. (in Russian)
2. Höppner F. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. Chichester: Wiley Intersciences. 1999. 289 p.
3. Viattchenin D.A. Control & Cybernetics. 2004. Vol. 33. P. 323-340.
4. Viattchenin D.A. Journal of Automation, Mobile Robotics and Intelligent Systems. 2007. Vol. 1. № 3. P. 29-38.

5. Viattchenin D.A. Journal of Information and Organizational Sciences. 2009. Vol. 33. P. 205-217.
6. Viattchenin D.A. Information Technology and Control. 2010. Vol. 39, № 4. P. 321-332.
7. Viattchenin D.A. Iskusstvennyj intellect. 2007. №.3. S. 205-216. (in Russian)
8. Anishchanka U.V. Bezopasnost' informatsionnyh technologij. 2012. № 1. S. 19-21. (in Russian)
9. Anishchanka U.V. Informatics. 2011. №.3. S. 98-111. (in Russian)
10. Viattchenin D.A. Journal of Computer Science and Control Systems. 2012. Vol.5, № 2. (to appear)
11. Viattchenin D.A. Acta Electrotechnica et Informatica. 2011. Vol.11, № 3. P. 42-50.
12. Viattchenin D.A. Iskusstvennyj intellect. 2011. №.3. S. 479-489. (in Russian)
13. Burillo P. Fuzzy Sets and Systems. 1996. Vol. 78. P. 305-316.
14. Viattchenin D.A. Journal of Uncertain Systems. 2009. Vol. 3. P. 64-80.
15. Grzegorzewski P. Fuzzy Sets and Systems. 2004. Vol. 148. P. 319-328.
16. Ju H. Fuzzy Information and Engineering. Berlin: Springer-Verlag. 2007. P. 875–883.
17. Fisenko V.K. Iskusstvennyj intellect. Intellectualnyje systemy. Donetsk: Nauka i osvita. 2011. S. 302-306. (in Russian)

RESUME

U.V. Anishchenko, D.A. Viattchenin, A.V. Damaratski, V.K. Fisenko

Detecting Anomalous Behavior via Discovering Attacks Systems for Interval-Valued Data

The paper deals with the problem of detecting anomalous user behavior in a distributed computational network. The data should be represented as a table where each cell of this table contains an interval of values.

The method of detecting outliers in interval-valued data was applied for the problem solving. The method is based on constructing stable clustering structure using a heuristic method of possibilistic clustering. Results of numerical experiment seem to be satisfactory. So, the proposed approach is effective tool for detecting anomalous user behavior in computational networks.

Статья поступила в редакцию 05.06.2012.